

McNamara Declaration

Exhibit 46

part 2

MIT Press became the first publisher to digitize-and-lend its backlist. The philanthropy Arcadia pledged \$1M to bring on other publishers.³

Meanwhile, our idea took hold at Boston Public Library, owners of the full trade collection of Houghton-Mifflin (HM). HM's Archivist convinced her president to digitize-and-lend their backlist, emphasizing in their MOU that "accessibility to all published literature is possible while still protecting the rights of authors, illustrators, designers and publishers."

In 2016, momentum continued with Kahle presenting "Transforming Our Libraries into Digital Libraries"⁴ to 90 leaders. We were invited to lunch with the new Librarian of Congress, Carla Hayden, seeding interest at the world's largest library.

We're now in dialogue with the Author's Guild to ensure them of the strict circulation controls this framework offers, seeking common ground. Remember that as recently as 2012, publishers weren't licensing any ebooks to libraries. The ALA played a pivotal role brokering consensus then, and we've asked their board to consider doing the same for digital lending now.

This year alone, IA teams have met with 1000 librarians, educators, Wikimedians, and legal experts. We've solicited feedback at DPLAFest, WikiCite, and the ALA Annual, where we received key endorsements from ALA and the PLA boards. (See Evidence.)

Media coverage continues to grow, including *EDUCAUSE Review*, *Library Journal*, and BBC. We're engaging the public through FacebookLive and YouTubeLive. Every week books from our wish list arrive, donated by Book Mooch members.

Public libraries are embracing *Open Libraries*, driven by their pressing needs for space and funds. You'll read letters from 35+ influencers and partners. After several years, *100&Change* has started a groundswell. Today we have the technology, buy-in from key leaders, and perhaps, with support from MacArthur, all the prerequisites for transformational change.

Stakeholder Influence

Through deep dialogue with learners, advocates for the blind, and engineers, we're discovering how better to serve the wide-ranging needs of people with disabilities. Since 2010, IA has been providing books in DAISY format and text-to-speech in our bookreader. But the community has told us that this is out-of-date. We now plan to improve accessibility of the navigation within books and produce more up-to-date formats, with consultation from LightHouse for the Blind and Benetech.

Our partners were clear about another point: reduce friction for their members. They're requesting one-touch authorization using their member databases, seamless integration with existing tools, and interoperability with the vendors they already employ. Many statewide library consortia plan to use New York Public Library's new mobile bookreader, *SimplyE*,¹ so we're integrating a seamless flow of our ebooks into this device. Fewer silos, more coordinated efforts.

We were surprised to learn that space is one of the biggest concerns for our early adopters. Many cities impose strict policies around weeding books, so libraries are renting expensive off-site storage for older titles. Storing books more compactly, while providing digital access to them will solve many problems for library budgets, space, and preservation. Counter to our assumptions, this suggests starting with the oldest post-1923 titles and working forward.

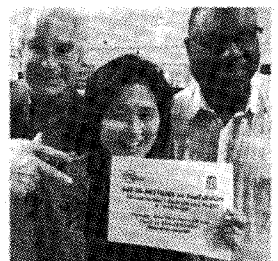
We approached the Digital Library Federation as a partner because of its explicit social justice mission. They have intentionally built a diverse membership of librarians who volunteer time to develop best practices around inclusion. It was DLF who suggested the “Inclusive Curation” methodology--a transparent, community-responsive process that we will be adopting. We’re learning from our partners at DLF everyday.

To truly serve global communities we need boots on the ground and community leaders at the table. With guidance from *Whose Knowledge?*, we are adding advisors in India, who are tying us to the on-the-ground networks with the most momentum.

In response to feedback from 20 top copyright experts, we are shifting our service plans and timelines. It makes sense to move swiftly to digitize and offer millions of the most useful ebooks to the print disabled, where copyright exemptions are clear-cut. Similarly, we can move forward with journalists and Wikipedians to ground their articles in citation “snippets.” We’ll increase efforts to buy ebooks directly from publishers and build partnerships with them akin to our understanding with Houghton-Mifflin-Harcourt. We’re also talking with the Author’s Guild directly.

But in response to the legal opinions we’ve heard, in Year 3 we may adjust our library circulation practices to start with those books that have negligible market impact: scholarly works, out-of-print works, and books filling in the missing century from 1924-on. Luckily, those are precisely the books that libraries want to transition off-site. Legal experts concur: we can clearly digitize any in-copyright book and share it with the print disabled. But we’ll work with all library partners to modulate circulation according to risk. In all these ways, our stakeholders are helping us to recalibrate our plans, honor their values, and address the concerns of creators.

C. Scalability of the Project



Planning for Scale

10% of the world experience disabilities that impact their ability to read.¹ For the blind, less than one in ten books are now accessible.² We have sustainable technologies to solve that. *Open Libraries* will scale across many dimensions, bringing more diverse books to millions of learners, in more accessible formats, to an increasing number of nations around the world. In the US, libraries offer free information services for all, but 80% of libraries are small and rural,³ where time-constraints and geography present barriers to access. We will address that by transforming the US public library system itself. In 2013, an average of 17% of their books were digital,⁴ by 2023, we'll make 80% digital. IA will ensure the long-term preservation of four million 20th-century books, now missing from our digital shelves. And we'll increase the number of ebooks available to the print disabled by 10x.

To achieve these goals, we must increase our technological capacity to scan books at scale while dramatically reducing costs, and also expanding the capacity of US libraries to lend their own ebooks.

Scaling Technologies

Starting with IA's collection of 2.5 million public domain ebooks, we plan to scale to approximately 6.5 million volumes. Scanning four million new books in five years requires increasing capacity at every juncture of our workflow—from tracking physical items to improving metadata and search. Since 2015, our engineering teams have been tackling every one of these areas. In 2014, IA released its next-generation scanner, the TableTop Scribe, featuring higher quality imaging and faster processing.⁵

Cost-effectiveness will be critical. Many libraries spend more than a dollar per page to digitize their materials. At our twenty North American scanning

centers, we charge 10 cents per page. IA estimates reducing that by 66% at our new super scanning centers in Hong Kong and Shenzhen.

IA has also refined our own data storage—at a fraction of the cost of Amazon cloud services.

Expanding Diversity

The Digital Library Federation (DLF) and Digital Public Library of America's (DPLA) Curation Corps of 40-50 librarian specialists will shape targeted lists of books from traditionally underrepresented communities. On a public website, community members can vet the books we select. Working with ALA's *Our Voices* Project, we will buy everything in their pipeline from minority publishers and authors.

We plan to digitize some of the nation's finest collections of multicultural children's literature and LGBTQIA collections from the San Francisco Public Library, focusing on the experiences of people of color, the disabled, and other marginalized groups. We'll focus on books in Spanish and Indian languages to better serve the print disabled in those regions.

Activating Library Networks

Launching with early archetype innovators of diverse library types, from public to academic, we will scale up through a group of key influencers who can move peer networks forward. We will activate large-scale adoption through centralized leverage points: vendors such as OCLC supporting 16,000 libraries,⁶ state-wide consortia, and certain state librarians who oversee public library systems for entire states.

The Credibility of the Proposed Solution:

In 2011, when IA launched its digital lending library, OpenLibrary.org, we recruited 100 libraries to contribute books for scanning and lending on their behalf. All 50 State Librarians endorsed the project.¹ These early partners form the network of innovators we can scale with again.

We **built an extensive digitization supply chain** for this pilot: IA receives free books from used-booksellers and libraries. We run 20 facilities, scanning books for 1,000 contributing institutions.

Today, **IA holds 2.5+ million ebooks and executes 1.5 million annual loans**. In the last six months, **43 million people** have used our online bookreader; **319.5 million learners have downloaded ebooks**. For six years, OpenLibrary has offered temporary digital access to copyrighted works in a manner respectful of creators, without major incident. We've proved that one library can succeed, but to scale we need to replicate this model with others.

In 2016, we launched a second-stage pilot with Georgetown University Law Library, our first archetype innovator. Today, Georgetown's community has the option to check out a physical or digital copy of a growing number of its books. Over the last few years, Georgetown's Michelle Wu, has presented her model to approximately 100 libraries, encouraging them to build "collaborative digital libraries" for the entire nation. Through her journal articles² and as a first mover, Wu is one of a dozen key influencers validating this legal path forward. She's part of a group of copyright experts drafting a joint statement on digital lending, establishing a legal basis for major adoption.

Support for the Change

At the recent American Library Association (ALA) Annual Conference in Chicago, 70 members of the Consortial Ebook Interest Group met to vent their “E-Grievances.” They listed the “silo thinking between print and e-;” the urgent need to aggregate data as an industry; and that “Librarians aren’t actually engaged in ebook issues—we just accept what we have been given.” Open Libraries addresses each of these pressing problems. It empowers US librarians to shape their own digital collections.

We’re enlisting a broad cadre of early adopters of every archetype, each representing an association of peers:

- Specialty: Georgetown Law Library, Washington DC
- Independent: Society Library, New York City, NY
- Mid-size college: Northeastern University, Boston, MA
- Large University: University of Alberta, Edmonton, Canada
- Small Public: Delaware County Public Library, Ohio
- Major Urban Public: San Francisco & Los Angeles Public Libraries, California

These widely trusted champions can influence their social networks toward adoption.

We’ve established an Advisory Board of 25 global influencers in education, law, technology, accessibility, and libraries, who mobilize constituents and legitimize our solution. Open Libraries is now endorsed by the Boards of the Public Library Association, ALA and the Chief Officers of State Library Agencies (COSLA), representing 50 State Librarians.

To mitigate legal opposition, we’re following the guidance of 20 copyright experts. A working group from Harvard, Duke, NYU, and Georgetown

Universities will publish a joint statement on the legal issues involved. This common understanding will reduce risk and speed adoption by clarifying the legal framework for university and state decision makers.

Advantage of the Proposed Solution to Alternative Solutions and to the Status Quo

In *BiblioTech: Why Libraries Matter More Than Ever in the Age of Google*,¹ John Palfrey writes, “People’s information habits have undergone a sea change—a major shift toward the digital...The problem is that libraries need to provide both physical materials and spaces as well as state-of-the-art digital access.”

Libraries haven’t met this demand, stymied by digitization costs, legal risks, and missing infrastructure. A look at IA’s own ebooks and Amazon’s physical holdings reveals a missing century of books between 1923 when public domain is clear-cut, and the digital era (**See Charts**). This status quo compromises research and learning.

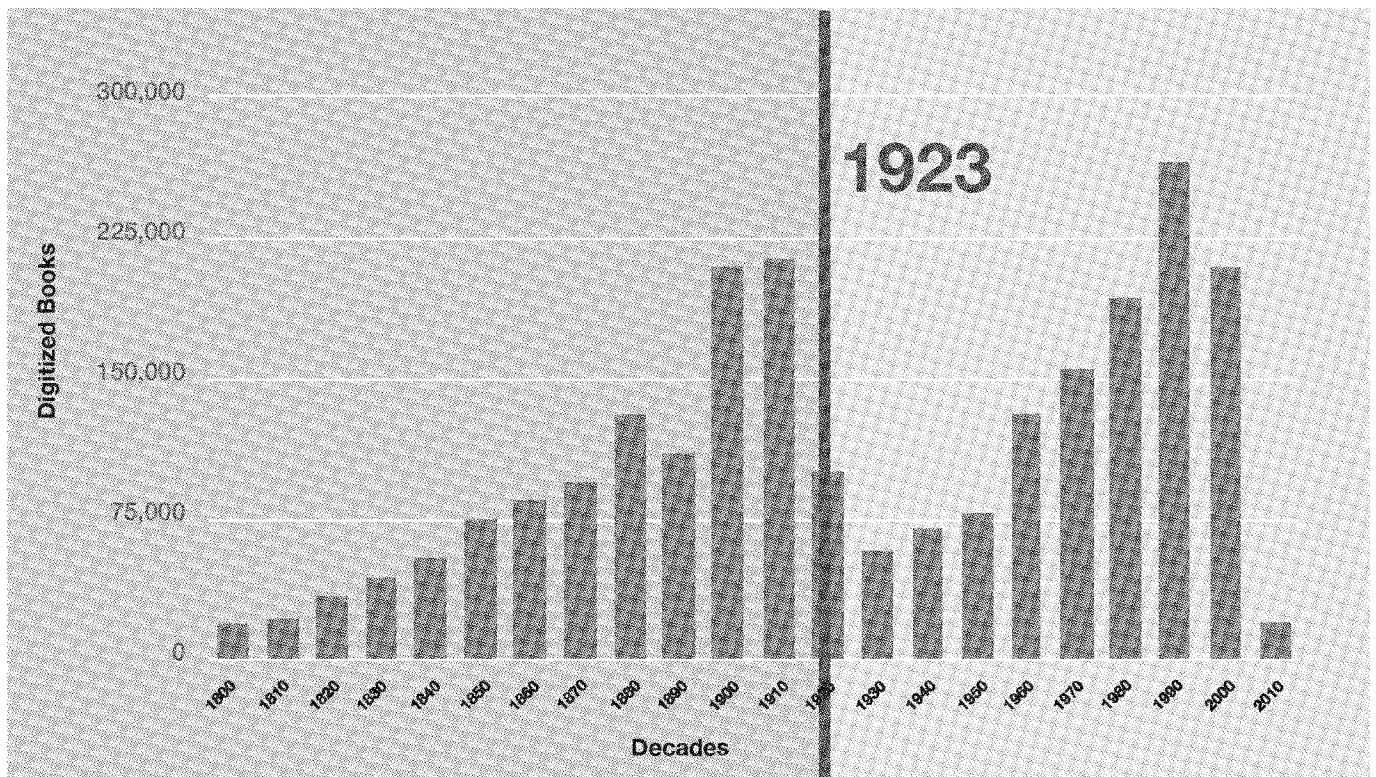
Google Books is the largest program to digitize millions of volumes. The resulting lawsuits sanctioned massive book digitization and access to snippets of in-copyright works. Meanwhile, as a for-profit company, Google makes no guarantees about reader privacy, preservation or long-term access.

Some Google scans are preserved in an academic repository called **HathiTrust**, where university subscribers can access public domain volumes. However, in-copyright works are not accessible even to paid subscribers. Publishers now license ebooks to libraries, but only very recent and popular books, often on temporary or restrictive terms.

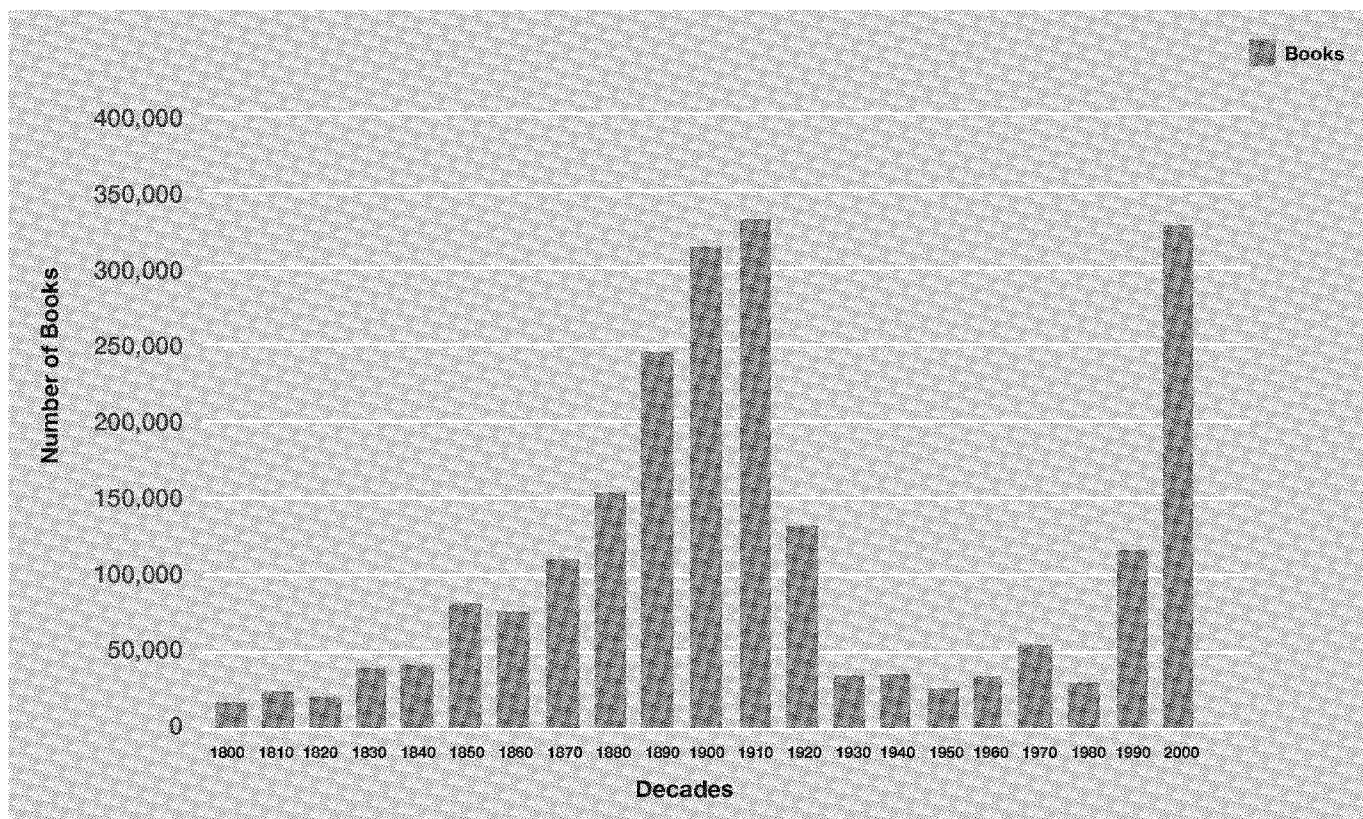
The growth of crowd-sourced sharing sites such as **Library Genesis** (3M

The growth of crowd-sourced sharing sites such as **Library Genesis** (3M ebooks) and **Sci-Hub** (52M articles) reflects a pent-up demand for easy, free access to books and journals. *Open Libraries* provides a legal path to some of that content. As MIT Press Director Amy Brand writes, “I see this effort as a proactive alternative to widespread circulation of unauthorized digital files.”

Internet Archive Digitized Books by Decade 1800 - 2010



Books in Amazon Warehouse by Decade 1800 - 2000



Taken from a random sample of 2500 new fiction titles available in the Amazon warehouse in 2012. Creator: Paul Heald, Professor of Law, Univ. Illinois Urbana-Champaign

Ease of Transferring and Applying the Solution at Scale

Over the past decade, IA has assembled the key elements to transition all US libraries to become digital libraries:

- Scanning capacity and a robust digitization supply chain,
- The legal foundation to support a Digitize and Lend Service (D&LS),
- Affordable, scalable technology,
- Demonstrable a model that can be replicated by other motivated institutions,
- Mobilized leaders in key constituencies to endorse the D&LS,
- A project plan to make a D&LS part of every US library's services.

The pathway to a D&LS at-scale features two work streams to broad, ubiquitous adoption. The first creates a large, compelling inventory of digital books. The second is a network of institutional advisors and service partners who will demonstrate how to launch the D&LS with minimal effort and proselytize the benefits to the community.

Why 4 million ebooks? We believe that a curated collection of 4 million titles, added to our 2.5 million public domain works, reaches an important threshold: our collection will overlap with the majority of most partners' collections, exerting the gravitational pull on later adopters to participate.

According to a 2013 survey of US public libraries, the average proportion of ebooks to physical books hovers around 17%. By analyzing data from 1200 library systems, our early results show we already hold 40% of the books in the average public library. With a more targeted inventory of 4M modern books, we can offer digital access to 80% of most library collections. These are the books that will overlap with college libraries and the millions of volumes in major public libraries.

Our greatest challenge is that libraries are heterogeneous—in type, technology, and funding-source—requiring multiple solutions. To scale, we will leverage centralized service providers and statewide networks that can activate large numbers through a centralized, single-point of contact. Perhaps our most important partner will be OCLC, the cooperative providing metadata infrastructure for the vast majority of US libraries (16,000 worldwide) and runs the cloud-based management systems for 600 academic libraries. Working with OCLC delivers two crucial capabilities—a widely available, easy-to-use matching service and the capability to surface those digital items in local catalog searches.

We're also integrating with mobile ebook platforms, including New York

Public Library's *SimplyE*, which is poised for adoption by hundreds of US libraries. *SimplyE* enables us instantly to offer highly motivated user groups with turnkey mobile access to their own ebooks.

IA has targeted the statewide consortia with the most advanced ebook strategies, including California, Connecticut, Georgia, Illinois, Minnesota and Wisconsin. At our webinar for State Librarians, the greatest interest came from rural states where library patrons travel great distances: Alaska, South Dakota, Tennessee, West Virginia and Florida.

We foresee a day when libraries everywhere will be expected to offer all their collections in digital formats. Amazon and Apple began that demand with the Kindle, iPad and iPhone, and patrons will demand it in their libraries, too. We offer the ecosystem an open infrastructure, high degrees of overlap, ease of transfer, and reduced risk.

Organizational Capacity to Implement the Solution at Scale

Since 2014, the Internet Archive has been engaged in organizational change and expansion. Our staff grew by 22% between 2014-16, adding five new directors who have the passion, expertise, and ability to sustain this project to fruition. We embrace Brewster's vision, and are capable of executing it.

Since 2014, we've been tweaking our engineering culture. It hasn't been easy—but we've evolved from a staff of brilliant individual contributors to one where skilled teams produce the most impact. During this project, we'll grow the number and size of teams along a path we've been forging for three years.

We're partnering with industry leaders, including OCLC and Benetech, that bring in-depth experience in their specialties. Benetech is the global leader in accessible digital content for people with disabilities. OCLC offers the

business and marketing teams that IA lacks.

Meanwhile, our evaluation partner, NewKnowledge, will be on hand to assist us in identifying roadblocks and adapting to new problems as they arise. We've intentionally structured the evaluation funds to be flexible and directed to outside experts in areas that may evolve.

There's something about the very DNA of IA's organizational culture that promotes scaling: we share deeply-held values and mission alignment; an open source ethos and a willingness to forego brand recognition as a provider of infrastructure. Librarians tend to be extremely collaborative, building upon the morals of our profession: reader privacy, commitment to sharing and open knowledge. All of these values will take us a long way toward scale.

Financial Sustainability of the Proposed Solution

MacArthur's support builds an enduring asset: a collection of 4M ebooks and the infrastructure to share them across an ecosystem.

To grow the collection, IA digitizes a steady influx of 250,000 books per year for paying partners.

To maintain the technology, we've identified three strong possibilities for sustainable income:

- One-time transfer fee for institutions that want to hold their own ebook files; this would allow them to build moneymaking services atop the collection.
- Activation fee when institutions "turn on" the ebook lending service. After early adopters turn on for free, we could charge a nominal fee for late adopters, with a sliding scale.
- Interlibrary Loan Digital Fulfillment: A 2013 survey shows public and university libraries paid between \$300-600 million to ship materials for interlibrary loans. OCLC runs an ILL matching service, executing 4.5 million ILL book requests



Open Libraries

a project by the Internet Archive

Internet Archive

100&Change

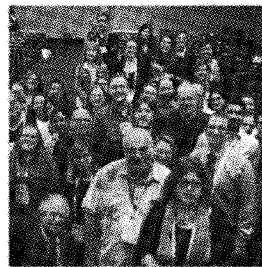
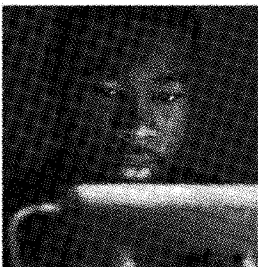
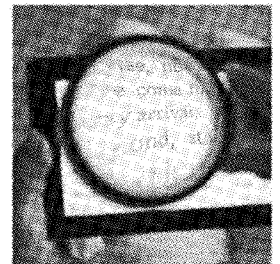
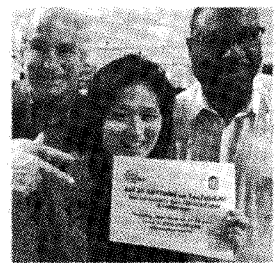
each year, averaging \$17. We can provide lower-cost ebook alternatives for ILL fulfillment. As the dominant ILL middleman, OCLC already has robust marketing, and service teams poised to expand to e-delivery.

In Year Three, IA and OCLC will conduct market research and build business plans for these three opportunities. In Year Four, we'll select the most promising and pilot them. In Year Five, we can start charging for some services.

Our plans expand the staff by 14% and reduce back to a 6% increase after Year Five. That growth seems sustainable with or without additional income.

When we are done, diverse communities will see their stories, now missing from library shelves, preserved forever.

D. Monitoring, Evaluation & Learning



MONITORING

Results

Open Libraries will transform the lives of millions of people around the globe by giving them unprecedented, free, long-term access to knowledge. This transformation will occur through one of society's most trusted institutions—the public library. Our goal is to bridge the informational divide that separates us along lines of income, geography, education, and physical ability. For many of the people who visit libraries 1.5 billion times each year,¹ the library has become their sole lifeline to trustworthy information—a place that is free of paywalls, ensures their privacy and preserves the stories that reflect their lives. Our solution combines scalable technologies developed by the Internet Archive, the information infrastructure anchored by libraries, and the membership networks of organizations with intimate knowledge of the needs of our most vulnerable populations.

Because *Open Libraries* has many facets, the monitoring and evaluation will be constructed around these key goals:

- Expanding global access to knowledge
- Increasing diversity in digital content to meet the needs of people from underserved racial, ethnic, gender and age populations, as well as those requiring special accommodations
- Expanding libraries' digital assets to maximize adoption nationwide and preserve their collections

To monitor progress in meeting these goals, we will focus on these outcomes:

- growth in digital assets made available to libraries and learners
- changes in usership patterns of digital assets compared to historical circulation
- expansion and use of assets for blind, low vision and dyslexic communities

- growth in titles and the digital content reflecting needs and interests of diverse communities
- growth in use and ebook availability to minority and underserved communities
- change in ratio of ebook vs. hard copy in library collections

The monitoring and evaluation strategy will be led by two teams with deep experience measuring change in the global library ecology. The first, **NewKnowledge Organization Ltd**, is a social science think tank that was recommended to us by the Gates Foundation's Global Libraries Program. With a decade of practice helping the Public Library Association (PLA) and the American Library Association (ALA) assess their impact, NewKnowledge is strategically placed to help align our efforts with other global library initiatives.

The second, are two research centers at the **University of Washington's (UW) Information School: the Technology and Social Change (TASCHA) group and the DataLab**, who will serve as the technical backbone of the project. TASCHA brings global experience measuring the impact of digital knowledge in the developing world; it's one of three "legacy partners" endowed with a decade of support by the Gates Foundation as Gates winds down its Global Libraries project.

Our evaluation teams will challenge our direction at key intervals, helping us to remain responsive, iterative, and true to our impact goals. Our monitoring program is designed as a transparent process tracking the "digital footprints" of our ebooks, how communities are using these materials, and their impact on the evolving ways libraries are helping communities bridge the informational divide.

Tracking

To track results, we're establishing two parallel initiatives. First, with TASCHA and DataLab, we'll create an **Impact Data Trust** to manage the quantitative measures of our project. The Trust will be a transparent, public repository for the ebook usage data that IA and others generate, held by institutions that maintain strict protocols for reader privacy and anonymity—something no for-profit analytics company guarantees. IA will preserve this data, but the UW teams will oversee its use, assisting researchers who want to query these results. By aggregating data from many sources, we hope to see the bigger picture, benefiting not only this project but the entire ecosystem. DataLab will develop the Trust's tools, protocols, and strategies to leverage these resources through its academic networks.

TASCHA and DataLab will issue data quarterly reports, using their teams to look for untold impact stories within our results. This work builds upon TASCHA's track record of investigating the impact of technologies in more than 50 countries, over a decade, including research on public libraries as venues for access to information.

The **Impact Data Trust** will capture the “digital footprint” of IA's growing ebook collection as it travels through libraries, Bookshare, and organizations for the blind to readers worldwide. Throughout the project, NewKnowledge will use an integrated set of quantitative measures and qualitative evaluation tools to provide a detailed understanding of the multiple impacts of this project for each beneficiary group.

Digital footprint data will tell us what content is being used and at what frequency. These data will point us in the direction of where to deploy targeted focus group studies. Working with Benetech and the Digital Library Federation, NewKnowledge will set up community panels to better

understand the information needs and priorities of different groups. Blind versus low-vision users, dyslexic learners, and ethnic communities all have unique ways of considering need and equity in library collections. DLF and Benetech's community feedback will complement IA's digital use metrics to tell a fuller story.

As questions arise about sub-populations arise, we'll respond to the evolving context. We've reserved 1.5% of the budget for specialized sub-contractors who will conduct surveys, focus groups, or other investigations to address new questions as they arise.

Meanwhile, NewKnowledge and TASCHA already monitor the library ecosystem, analyzing data to help increase libraries' impact at scale. We can leverage their wider work to add to our own, including resources from: the International Federation of Library Associations, the Public Library Association, Benetech, the Digital Library Federation, the Gates Foundation-supported *Libraries Transforming Communities* and the *National Impact of Library Public Programs Assessment*. The Internet Archive's data scientist, Vinay Goel, has already found insights by crossing the collections of 1200 library systems with our own book collections. We can only imagine what he will find once we establish the Impact Data Trust.

It's exciting to imagine IA engineers, librarians, social scientists, data scientists and community panelists working together to analyze what it means for people when they have free and unfettered access to more and better information.

Milestones:

Year 1: Building Infrastructure: Influencers, Validators, and Innovation Partners

Year One rolls out in three phases: A) building the human and technical infrastructure; B) securing outside validation; and C) engaging the first wave of partners for testing & community feedback loops.

- Core Staff hired, including Key Books Leaders
- Advisory Board meets in 7 Working Groups
- Legal scholars build consensus around fair use arguments
- Create College Core/Public Library Core Book Lists
- Purchase ebooks (from publishers, authors)
- Scanning Center capacity—174,000 books
- Engage Archetype Innovators in first-wave testing
- Establish the Promise: Media Campaign “Everyone Deserves to Learn”
- Secure key endorsements
- DLF launches Inclusive Curation Project & dashboard for public feedback
- M&E: assumption testing, goal setting, forming sampling frames, and feedback panels
- TASCHA establishes Impact Data Trust

Year 2: Refining Technologies & Setting Priorities

In 2019, we focus on A) refining our technologies; B) building a compelling collection of books; and C) expanding through established library networks.

- DPLA Curation Corps select books for diversity and education
- Establish “Most Widely Held” and “People with Disabilities” lists

- IA improves accessibility of site & books
- Scanning Center capacity—198,000 books
- Books flowing in from many sources (booksellers, paying partners, libraries)
- Scaling through statewide library networks and service partners
- Raising Awareness: Media Campaign “Inspired to Lead”
- M&E: diversity feedback and priority setting among collaborators
- TASCHA: first data subcontracts issued for outside learning
- Financial modelling: Three Business Options with OCLC

Year 3: Expanding Reach, Reassessing Impact, & Gap Analysis

In 2020, IA builds a compelling collection & circulation system. We expect to see network effects in both reach and impact.

- Large user groups engaged: Wikipedians, OER Educators
- Certified People with Disabilities now authorized for ebook delivery
- Global expansion towards Marrakesh Treaty nations; Benetech launches Spanish Portal
- Scanning Center capacity—723,000 books
- Circulation & Integration systems built
- IA Partner Support Teams conduct training, onboarding
- Media Campaign to engage public participation
- M&E: reassess impact, recalibrate priorities, gap analysis
- TASCHA: administering Impact Data Trust research for data-driven insights
- Business model trial

Year 4: Recalibrating for the Final Sprint

In 2021 our focus shifts to A) content discovery and B) understanding impact. We have a final chance to recalibrate and fill gaps.

- Discovery through Wikipedia citations & direct links
- Close to turning 80% of the average public library collection digital
- Global expansion towards Marrakesh Treaty nations
- Scanning Center(s) capacity: 742,500 books
- Overlap with majority of major public libraries' collections
- Partner Support team builds self-guided tutorials
- M&E Team: final priority setting and recalibration
- Business Model validation—service ramp up

Year 5: Legacy Planning for Infrastructure & Support

In 2022, we focus on building the legacy infrastructure and institutional support to sustain and grow Open Libraries' assets.

- Staff ramp down as key infrastructure is built and stable
- Scanning Center(s) capacity: 742,000 books
- Ramp up: paying scanning partners in super centers
- Major overlap with mid-size college collections & books cited in Wikipedia
- Late adopters join
- M&E: retrospective impact analysis
- TASCHA: support planning for Impact Data Trust
- Decisions re: business for legacy support

EVALUATIONS

Changes for Beneficiaries:

For the purposes of evaluation, we divide beneficiaries into five categories:

- **Learners who both seek and need greater access** to books in digital format, including people living far from a library, housebound, elderly, or those physically unable to get to a library to borrow physical materials. They may also be people who increasingly rely on and use digital materials in their daily lives.
- **Individuals who require physical accommodations** to benefit from print resources, including blind, low vision, mobility-restricted, and dyslexic learners, and others who require digital tools to access learning content.
- **Groups excluded by the informational divide**, including minority communities and lower-income Americans, many of whom rely on library services to fulfill their information-driven needs.
- **Diverse communities** whose cultures are underrepresented in library collections.
- **Library professionals** in the midst of a significant transformation of services they provide.

One of the reasons we chose NewKnowledge is their impressive track record designing evaluation programs that recognize the need to work directly with those most at-risk. They recognize that data representing majority audiences can conceal the needs of minority groups. Thus, NewKnowledge's strategy is built upon qualitative feedback from a network of community-based panels. Throughout the monitoring process, NewKnowledge will measure not only how many titles we've added or what kind of content we've made available, but they'll work with all of these communities, ensuring each addition reflects their needs. They will report to us each quarter, informing us if the data is representative of the lived experience of library patrons. These evaluation strategies bring new communities into the planning process over our five-year project.

NewKnowledge will recruit representatives from each beneficiary group in waves over the five years for qualitative research. This includes community panels of users from partner libraries—coordinated by leaders from the Digital Library Federation. Working with their cultural assessment model and expert knowledge of each cultural group’s needs, DLF works to develop a community voice that helps set priorities for book selection and independent cultural review of how needs are being met. TASCHA will issue quarterly benchmark updates to support review and priority-setting for the work’s next phase.

As part of a consolidated needs report, NewKnowledge will undertake regular community-wide surveys, benchmark reviews from across the library field, including both early adopter sites and user interviews. They will work with Benetech to evaluate how their 486,000 members are using and benefitting from access to 10x the books they had before. Perhaps more importantly, they will gather data from libraries who have not adopted the Digitize and Lend Service (D&LS) and people with print disabilities who are not downloading a broader range of books, to learn why some of these areas have stalled. These reports will highlight how *Open Libraries* is removing barriers to knowledge, which books have multiplier value across sub-groups, opportunities for expansion, and the status of persistent barriers to adoption. If we aren’t hitting scale or reaching people, NewKnowledge will work with us to discover why.

Data Sources

The Internet Archive sits at the nexus of three data-driven professions: technology organizations with our ability to gather data dynamically at scale; libraries, creators of the union catalog of books featuring the best structured metadata of any profession; and the Open Knowledge world of Wikimedia Foundation, dedicated to posting all its data in an open, transparent fashion. We will be able to gather data from all these sources—most likely lots of data. But there's a difference between gathering data and making sense of it.

That's why our strategy includes the data scholars at University of Washington's DataLab. UW is renowned for its cross-disciplinary research, pairing social scientists like TASCHA's Chris Coward with data scientists like Dr. Carole Palmer. It will take the stellar reputation of the UW Information School to convince libraries to entrust their readers' data to an outside institution. But once convinced, we have the opportunity to gather data from the hundreds of libraries using NYPL's mobile e-reader *SimplyE*, data about the 9 million interlibrary loan exchanges brokered each year by OCLC, not to mention the circulation data of thousands of partner libraries. These data will tell us which library's patrons are borrowing the books and other key use data. Using sampling metrics, NewKnowledge will be able to extrapolate understanding about who is reading. But understanding why these results are important will take a mix of quantitative and qualitative methods.

Mixed-method evaluation relies on the careful arrangement of quantitative data and parallel qualitative data that can explain potential variations in results. Surveys, interviews, and large scale user and non-user data will inform this project. For *Open Libraries*, these data will come from an array of sources, including IA digital activity, data aggregated by project partners, circulation data about physical books from a stratified targeted sample of

partners, and formative evaluation studies of users vs. non-users.

Big data alone does not always tell an accurate story of impact. Given our wide array of beneficiaries, our methods require validation of any quantitative data for its truthfulness to each community. To assure validity, we'll use benchmarking templates for each service community. DLF will help us convene community panels, issue surveys of their working groups, and gather public feedback via their dashboard. NewKnowledge will conduct case study interviews exploring how the service is evolving, with everyone from Chief Technology Officers to end users.

Our plans call for TASCHA to lead a yearly "Data Mobilization Discussion" to review the data, consolidate feedback, and conduct case study interviews with our peers about how the practice is evolving. We will award small sub-contracting grants to specialists who want to dig deeper into these troves of data.

Methodology

Our evaluation team is committed to **"Mixed Methods" evaluation**, a strategy that recognizes the importance of collecting both quantitative and qualitative measures to gain a full picture of impact. This approach will assess the impacts of the project for each group of beneficiaries listed earlier in this document.

Evaluation studies will collect data from each beneficiary community over the life of the project, studying patterns and needs as they emerge. Working with each partner, the evaluation team will monitor how new titles are selected and how they represent the voice of the communities they serve. It will track adoption rates among libraries and identify challenges as well as successes as the process continues. It will validate the growth in digital collections and

document their usage. It will further look at the effectiveness and learnings among the many project partners.

We will draw data samples about usage and collection growth from a wide range of public, academic, and specialist libraries, urban and large libraries, rural or small libraries, stand-alone destination libraries, early adopters and those who remain resistant or minimally involved. To assure the growth of materials reflects the needs and interests of diverse communities, we anticipate using public forums, open discussion groups, spot surveys and open response tools to gather feedback throughout the project. New data will be benchmarked against other information residing in the DataLab, the research group residing at the UW Information School. Methodologies also include the organization of annual “Data Mobilization” summits, bringing together partners to review activities and findings, as well as publications to share with all partners.

All data will become part of the **Impact Data Trust**, developed and managed by TASCHA. This information will not only be useful to help us develop the best line of inquiry and data gathering as we move forward, it will also serve the larger needs of the field. We anticipate new avenues of inquiry to arise throughout this project. We will therefore support sub-contractors with an interest in pursuing relevant questions, designed to add continually to the understanding necessary for the overall project to be effective.

The expertise of NewKnowledge, TASCHA, and DataLab will continually inform the evaluation process, determining the range and type of relevant methodologies, recognizing new opportunities for pertinent data collection and research, and being fully responsive should a change in direction be warranted.

Our measurement and evaluation strategy—focusing on regular updates, rapid feedback reports, and teams that collaborate closely on data implications—outlines an approach that will allow us to make continual course corrections to ensure our work is equitably serving those most in need, identify paths to broaden or deepen our impact as the program evolves, and ensure we are able to leverage resources as ebooks shift the access landscape.

Gender Analysis

Open Libraries seeks to provide a more equitable distribution of assets to all members of society. For girls with disabilities, the divide is even wider; fewer attend school than boys, exacerbated by a lack of accessible books. *Open Libraries* addresses this issue, particularly in India, where we'll partner with Centre for Internet and Society to address girls' education needs. In June, the Digital Library of India contributed 463,000 Indian ebooks to IA (50 languages). By analyzing this collection with a gender frame, we'll prioritize improving the quality of certain ebooks to achieve the greatest impact for girls.

We'll also focus on balancing representation of women in digital collections. As Advisor, Anasuya Sengupta, writes:²

When a young Dalit woman from Daulatabad with no easy access to libraries cannot read her own community's scholarship, she cannot fight back the oppression of centuries of being considered "untouchable," and having her history and knowledge denied...The reality is that if you're not online, you're not in Wikipedia, you don't exist.

Today, there are few histories written by or about women and even fewer online. To redress this, our evaluation prioritizes equity as a measure of collections development, to find female authors from the early 20th century and those representing the history of non-binary gender representation. In his paper *M or F*, our evaluator John Fraser explored fallacies that accrue from using binary data analysis to report on gender impacts. These critical questions are where our qualitative evaluation will ensure IA's approach is inclusive and accurate.

LEARNING

Findings

To quote Chicago Public Library's Commissioner, Brian Bannon, "Libraries always have been, and continue to be, in the knowledge business, not "static or unchanging institutions. Instead, libraries have evolved to support knowledge creation and dissemination in ways that are relevant to our communities' changing needs." One focus of our work, then, is to accelerate this evolution.

Open Libraries is, by nature, a learning endeavor, working iteratively with the library community to shape greater understanding of how digital access can liberate all learners.

Our learning outcomes include:

- Digital Infrastructure innovation, an enduring public legacy of this investment.
- Process innovation that moves libraries' decision-making about collections from behind-the-scenes to a national dialogue on the needs of the underserved.
- Collection Development innovation based on analysis of aggregated data

about circulation.

- Inclusive Curation Innovation from the community panels and working groups at DLF.
- Data-Driven research from TASCHA that can guide strategic library growth.
- Learning about the aspirations of print disabled communities who have traditionally been offered a minimum of services rather than the full range of opportunities.
- Shifting perceptions of value encouraging long-term preservation of collections for future generations.
- Understanding changing space allocations that may expand opportunities for new library learning programs.
- Expanded legal understanding of the necessary strategies for public institutions to serve their online, networked communities.
- Identifying how readily available, trustworthy information can combat the trend toward “alternative facts” that undermines democratic reasoning and decision-making.

These learnings will impact how we set our priorities, allocate our budgets, and rebalance resources to address problems or accelerate successes. We’ll monitor different signals with our legal advisors to adjust our plans according to perceived risks. In India, we’ll have on-the-ground advisors to monitor how 6.5M books are reaching people with print disabilities and sub-contractors to investigate any “last-mile” issues.

But data, impact reports, colloquia—these only go so far in understanding a dynamic system. So we’ll go out in the field, hang out in youth medialabs, and talk to librarians and leaders to get a holistic picture of the systems we’re helping to change.

Two years ago, we led a summer reading program for children of Internet Archive staff. Each week we’d digitize an award-winning book and come together to discuss it. A few children came and went, but the program was really set up for Eileen Alfaro, a bright 4th-grader whose mother Roxana works with us. It became apparent from our conversations though, that

Eileen wasn't finishing the reading. She had a computer, we digitized the books—*was this straight-A student slacking off?* We came to understand the problem was the poor Wi-Fi in Eileen's apartment that made it almost impossible for her to flip the pages in our browser-based bookreader. Eileen needed to *borrow* the book, to download it for a 2-week span, in order to read successfully.

As we're building this global service, we remember Eileen. We'll take the time to meet others like her, to make sure all the signals line up with the lived experiences of those we serve.

Dissemination

As outlined in the Project Plan, *Open Libraries* is an open, transparent, community-responsive project that has scheduled milestones and staff dedicated to sharing our findings.

Strategies include

- Annual Data Mobilization Workshop (UW Information School)
- Annual Library Leaders Forum
- Annual IA Celebration with 700 supporters
- Two-way communication through DLF's Dashboard
- Public dissemination through IA's blog, Facebook page (118,000 followers), newsletter (circulation: 750,000), OpenLibraries.online website
- Legal presentations at Fair Use Week events
- Wikimedia's public documentation and engagement at Wikimania, Wikicite

Our **M&E Teams and advisors**—who include nationally recognized experts in law, education, technology and library sciences—will disseminate findings



Open Libraries

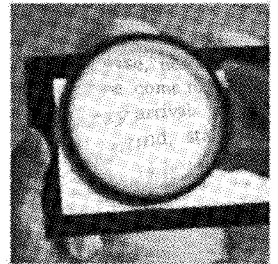
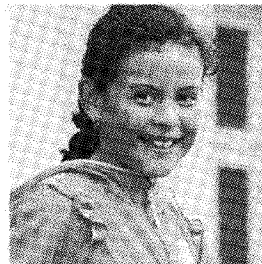
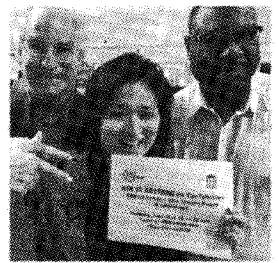
a project by the Internet Archive

Internet Archive

100&Change

through academic articles and professional circles. **Brewster Kahle** is a frequent keynote speaker. *Open Libraries* Team members will present our findings at dozens of conferences and fora. Through all these channels we will share our lessons with decision makers in law, education, local government, libraries and philanthropy to amplify impact.

E. Additional Information



Detailed Budget Narrative

This five-year budget reflects **expenses totaling \$101,142,290**, offset by **\$1,143,096 in project income, including an \$1 million Arcadia grant** to digitize publisher's backlists in 2018-2019.

Largest budget categories: acquisition/shipping of physical and digital books (13% of budget); scanning and storing those books (46%); and staff and fringe (23%). These figures assume:

- Since 2006, IA has scanned approximately 2.5 million public domain books and 540,000 modern works.
- In our 4M book lending collection, we assume 825,000 books (20%) of our curated list will come from our existing collections or those uploaded by others. For instance, the Digital Library of India recently contributed 465,000 books in 50 languages.
- 585,000 books will be added by libraries that pay IA to scan their books. (Since 2004, IA has digitized approximately 250,000+ books per year for paying partners.)
- 10,000 ebooks will be purchased directly from publishers/authors. This is a modest figure, since there is no current pathway for libraries to purchase ebooks rather than license them.
- 2,580,000 books will be purchased or donated to IA, scanned, and physically archived in our repository.

Our **long-term storage rate** of \$5/book reflects the cost to store a physical book in our existing repositories. These warehouses are owned by a separate non-profit entity, The Open Library of Richmond, which charges IA below market rates.

Digital storage and serving cost is \$2.00 per book, representing an average 1.5 gigabytes for a typical high resolution set of files. We base this on **Stanford researcher, David Rosenthal's** study of our long-term data center costs. This is significantly lower than commercial cloud storage rates.

Approximately 4% goes to **Benetech**, to help us increase the accessibility

**Open Libraries**

a project by the Internet Archive

Internet Archive**100&Change**

of our texts and provide authorization for its 475,000 members to download any book on archive.org. Benetech will also develop a new platform serving people with print disabilities in the 12 Spanish-speaking Marrakesh Treaty signatories, 10 of which do not have any national ebook platform for people with disabilities.

We devote 4% to measurement and evaluation activities with outside teams, **NewKnowledge Organization**, and **University of Washington's Technology and Social Change (TASCHA) Research Center**, which will oversee an "Impact Data Trust" for libraries, an open repository for user data from IA and others. We allocate flexible funds for expert consultants to use this data to analyze our impacts and system-wide effects.

In Year 3, we'll research three possible lines of business, construct business plans for the most promising and establish a pilot by Year 4. For instance, OCLC will research the potential for integrating IA's electronic lending library with OCLC's Inter-library Loan (ILL) system. In 2013, libraries spent approximately \$300M for ILL fulfillment, so we see tremendous market potential to provide e-delivery of those same books. However, this 5-year plan is not dependent upon earned income.

IA reserves 6.8% of the budget for indirect expenses supporting this work.

Your support creates an enduring information asset and the financially sustainable infrastructure to access it globally.